AD-A103 852 WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER F/G 12/1
AN EXAMPLE OF THE USE OF ANDREWS' PLOTS TO DETECT TIME VARIATIO--ETC(U)
JUL 81 A M HERZBERG DAAG29-80-C-0041
UNCLASSIFIED MRC-TSR-2239 NL

END
DATE
FILMED
0 8
DTIC

LE...

AD A103852

MRC Technical Summary Report #2239

AN EXAMPLE OF THE USE OF ANDREWS' PLOTS TO DETECT TIME VARIATIONS IN MODEL PARAMETERS AND OUTLYING OBSERVATIONS

Agnes M. Herzberg

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

July 1981

(Received June 23, 1981)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

81 9 08 040

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

AN EXAMPLE OF THE USE OF ANDREWS' PLOTS TO DETECT TIME VARIATIONS IN MODEL PARAMETERS AND OUTLYING OBSERVATIONS

Agnes M. Herzberg†

Technical Summary Report #2239
July 1981

ABSTRACT

Andrews (1972) introduced a method of plotting high-dimensional data in two dimensions. This method is exploited as a graphical tool for the examination of changes over time in the parameters of a time series model. An example using a Fourier series model is given to illustrate the method. It is also shown how outlying observations in the data can be found.

AMS (MOS) Subject Classifications: 62M10, 62H30

Key Words: Andrews' plots, time series, outliers, spurious observations, exploratory analysis.

Work Unit Number 4 (Statistics and Probability)

---

†Imperial College of Science and Technology, University of London, London, U. K.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

Andrews (1972) introduced a method of plotting high-dimensional data in two dimensions. In his method, Andrews represents each multidimensional point by a Fourier function. The clustering of plots of these functions is equivalent to the clustering of the multidimensional points. Andrews' method is exploited as a graphical tool for exploratory data analysis for the examination of changes over time in the parameters of a time series model. An example using the total Canadian unemployment figures from 1956-1975 is used to illustrate the method. These data have four spurious (outlying) observations and it is shown how these may be detected by the use of Andrews' plots.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

AN EXAMPLE OF THE USE OF ANDREWS' PLOTS TO DETECT TIME VARIATIONS IN MODEL PARAMETERS AND OUTLYING OBSERVATIONS

Agnes M. Herzberg[†]

1. Introduction

A graphical method is given for the examination of changes over time in the parameters of a time series model. This method can be used as an aid in exploratory data analysis. In a previous paper, Herzberg and Hickie (1981), the method is presented and two examples are given. A brief description of various multivariate graphical clustering methods and the use of Andrews' plots as a graphical tool in time series analysis is also given in Herzberg (1981). Here another model is used with a different set of data and further discussion given of the detection of outliers, or spurious observations.

2. Andrews' Plots

Andrews (1972) proposed the following simple and useful method of plotting high-dimensional data in two dimensions. If the data are m-dimensional, each point $\underset{\sim}{x} = (x_1,\ldots,x_m)$, where x_i $(i = 1,\ldots,m)$ are the measured variables, is represented by the function

$$f_{\underset{\sim}{x}}(t) = x_1 \cdot 2^{-\frac{1}{2}} + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + x_5 \cos 2t + \ldots \qquad (1)$$

plotted over the range $-\pi < t < \pi$. The functions given by (1) have several properties including the preservation of means, distances and variances and will also give one-dimensional projections. Thus, when (1) is plotted for

[†]Imperial College of Science and Technology, University of London, London, U. K.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

each data point $\underset{\sim}{x}$, the clustering of the points may be seen by a banding together of the plots of the functions. Tests of significance may also be made; see Herzberg and Hickie (1981).

3. Variation of Model Parameters

Herzberg and Hickie (1981) considered the regression model

$$\underset{\sim}{Y}_j = \underset{\sim}{X}\underset{\sim}{\beta}_j + \underset{\sim}{U}_j \quad (j = 1,\ldots,T-n+1),$$

where T is the total number of observations, n is the number of observations in each subgroup of observations used for estimating the unknown parameters, $\underset{\sim}{Y}_j = (y_{1j},\ldots,y_{nj})'$ is a $n \times 1$ vector, y_{ij} being the ith observation in the jth subgroup $(i = 1,\ldots,n)$, $\underset{\sim}{X}$ is the $n \times m$ matrix of the regressors, $\underset{\sim}{\beta}_j$ is the $m \times 1$ vector of unknown parameters to be estimated by least squares and $\underset{\sim}{U}_j$ is the $n \times 1$ vector of error terms. All the elements of the $\underset{\sim}{U}_j$'s are assumed to be independent and normally distributed with mean 0 and variance σ^2. It is assumed that the T observations are taken sequentially over time and it is desired to examine the variation in the $\underset{\sim}{\beta}_j$ over time.

Let $\hat{\underset{\sim}{\beta}}_j = (\hat{\beta}_{1j},\ldots,\hat{\beta}_{mj})'$ be the $m \times 1$ vector of least squares estimates of the elements of $\underset{\sim}{\beta}_j$ obtained from the jth set of n observations $(n < T)$, i.e. $\hat{\underset{\sim}{\beta}}_1$ is estimated from the first n observations, $\hat{\underset{\sim}{\beta}}_2$ is estimated from the second observation to the $(n + 1)$st observation, etc. From each $\hat{\underset{\sim}{\beta}}_j$, a plot of the function $f_{\hat{\underset{\sim}{\beta}}_j}(t)$, defined in (1), over the range $-\pi < t < \pi$ was made. The plots of these functions will show the change over time in the vector of coefficients $\hat{\underset{\sim}{\beta}}_j$. The plots, $f_{\hat{\underset{\sim}{\beta}}_j}(t)$, can be considered as a graphical weighted moving average. For each t a different weighting is given to the observations.

4. <u>An Example</u>

Table 1 shows the total Canadian unemployment figures from January 1956 to December 1975. It can be seen that the values for January 1958, 1961, 1971 and 1975 could be considered as being outliers or spurious observations in the data. Figure 1 gives a plot of these data.

The model

$$E(y_{j+i-1}) = \beta_{1j} + \beta_{2j}\sin\frac{2\pi i}{12} + \beta_{3j}\cos\frac{2\pi i}{12} + \beta_{4j}\sin\frac{4\pi i}{12} + \beta_{5j}\cos\frac{4\pi i}{12} \quad (i = 1,\ldots,12;\ j = 1,\ldots,229)\ , \tag{2}$$

where y_{j+i-1} is the observed unemployment figure in month $j+i-1$, was fitted to the data by least squares for each j fixed and $\hat{\beta}_j = (\beta_{1j},\beta_{2j},\beta_{3j},\beta_{4j},\beta_{5j})'$, the least squares estimate of β_j obtained. The plots of the function

$$f_{\hat{\beta}_j}(t) = \hat{\beta}_{1j}\cdot 2^{-1/2} + \hat{\beta}_{2j}\cos t + \hat{\beta}_{3j}\sin t + \hat{\beta}_{4j}\cos 2t + \hat{\beta}_{5j}\sin 2t \quad (j = 1,\ldots,229)\ , \tag{3}$$

were obtained and plotted. Note that (3) differs from (1) but the mathematical properties of (1) are retained. Several variations of (1) were tried but the outlying plots were most easily seen when (3) was used. This is due to the particular weighting which (3) gives to the $\hat{\beta}_{ij}$'s. and thus to the individual observations.

It could be seen from the plots when plotted in chronological order on a graphics terminal that certain ones stood out from the others. Any long term increases or decreases in the plots were also noted.

The 229 Andrews' plots are given in Fig. 2 and Fig. 3. The plots in Fig. 2.k $(k = 1,\ldots,12)$ are those obtained from (3) for $j = k,\ k + 12,\ k + 24,\ldots,k + 108$. The plots in Fig. 2.k are similar except for the ones

Table 1. Total Canadian unemployment figures for 1956-1975 in 1000's.

	January	February	March	April	May	June	July	August	September	October	November	December
1956	315	340	321	273	175	127	112	116	116	110	149	211
1957	328	352	378	334	209	177	181	194	214	223	318	422
1958	579	600	638	553	388	339	310	317	284	328	378	466
1959	577	570	553	466	354	248	239	257	224	250	316	405
1960	546	597	607	550	417	313	328	350	325	366	426	525
1961	690	716	702	619	454	367	351	320	305	315	347	411
1962	543	582	559	484	335	300	307	279	259	282	342	414
1963	541	546	550	463	347	305	294	271	251	266	303	346
1964	466	467	456	403	293	282	265	246	217	257	257	284
1965	407	397	387	371	265	257	244	211	176	171	220	252
1966	359	356	341	298	247	230	244	228	205	195	238	266
1967	381	396	400	365	304	292	284	247	219	254	289	353
1968	464	482	488	436	366	395	371	319	262	288	338	373
1969	467	473	448	432	386	383	349	318	279	314	354	383
1970	485	526	542	544	513	529	518	548	398	419	476	538
1971	668	675	650	659	543	551	514	455	434	447	503	530
1972	665	627	642	592	552	568	543	503	459	483	524	584
1973	688	655	608	570	493	503	461	433	421	429	468	512
1974	637	635	599	568	524	469	465	447	431	430	493	597
1975	817	839	840	795	714	704	653	623	586	576	640	697

FIG. 1. Canadian monthly total unemployment figures from 1956-1975 in 1000's.

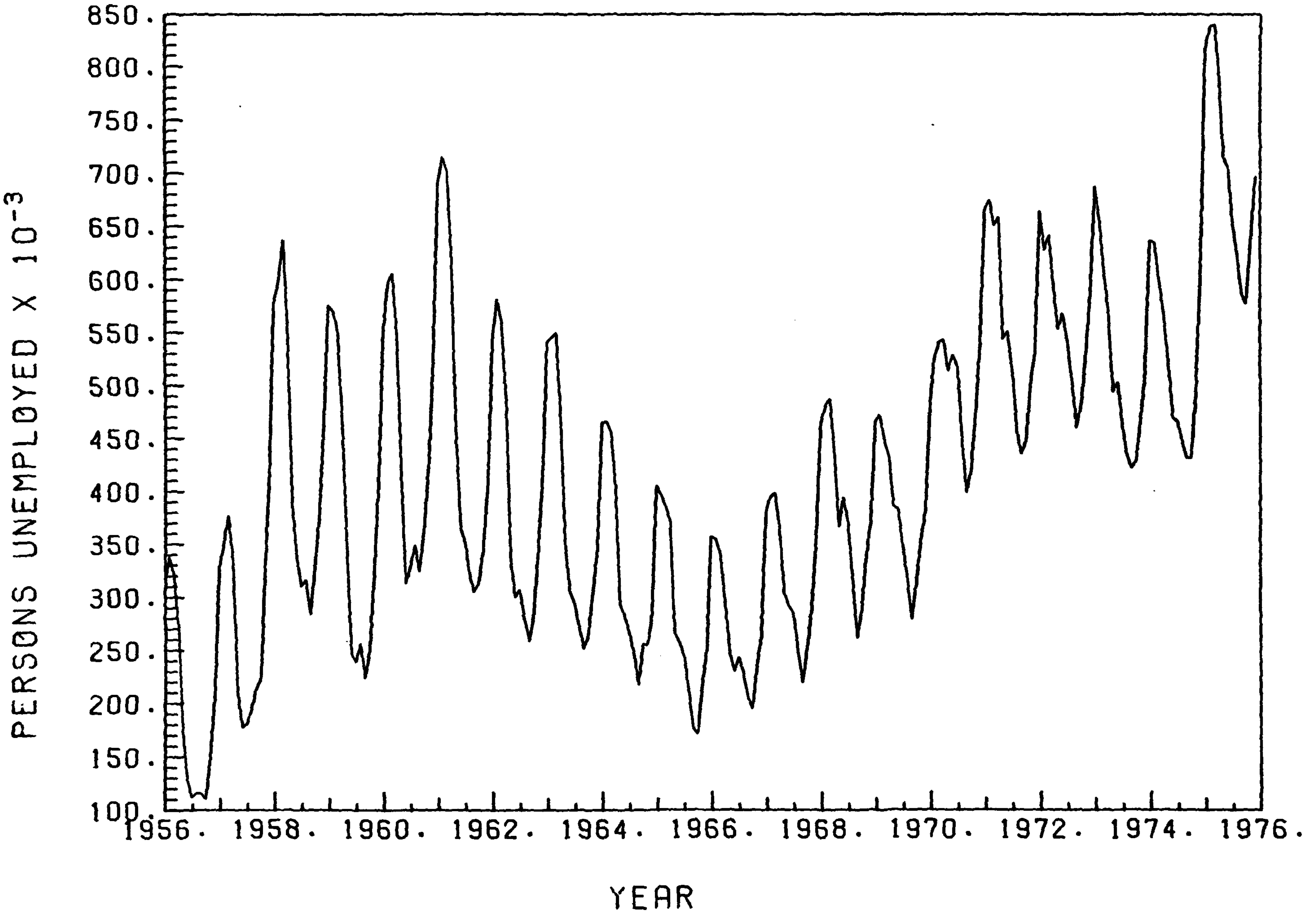


FIG. 2. Andrews' plots, $f_{\hat{\underset{\sim}{\beta}}_j}(t)$ $(j = 1,\ldots,120)$ given by (3), $\hat{\underset{\sim}{\beta}}_j$ obtained from (2) and sorted;

Fig. 2.k $(k = 1,\ldots,12)$ consists of plots for $j = k,\ k + 12,\ k + 24,\ldots,k + 108$.
(The darker curves denote these plots obtained in part from January 1958 or 1961.)

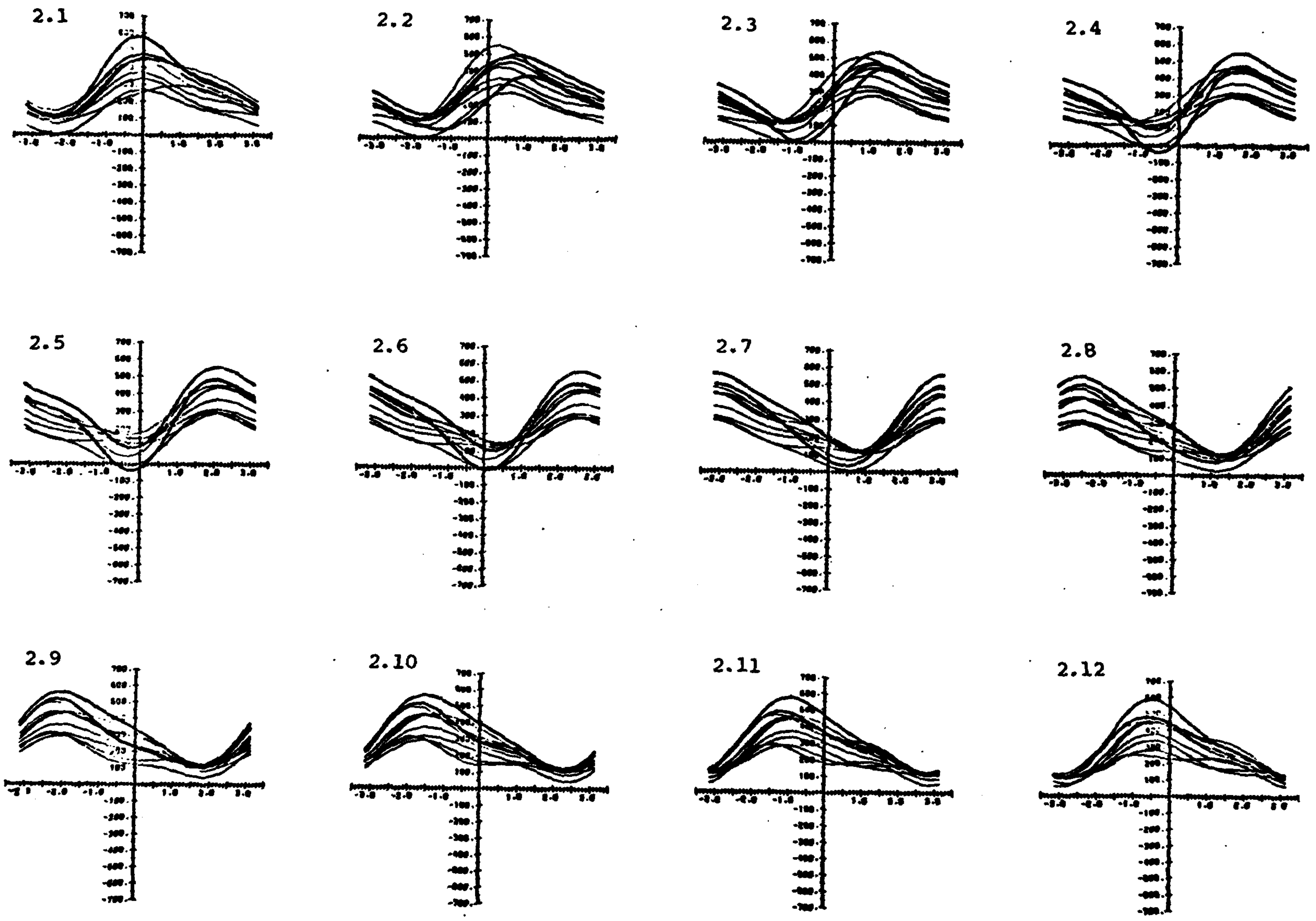
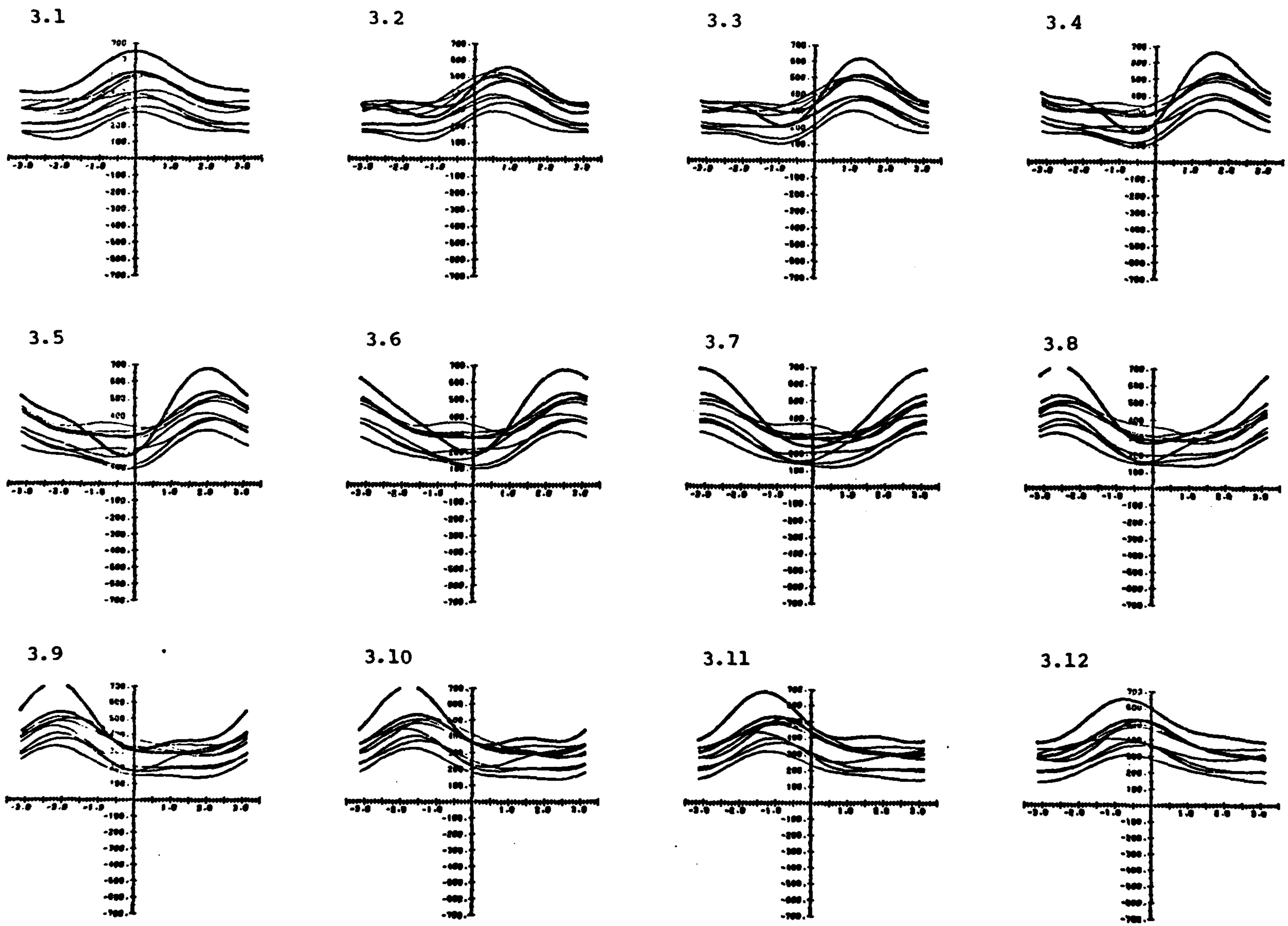


FIG. 3. Andrews' plots, $f_{\hat{\underset{\sim}{\beta}}_j}(t)$ $(j = 121,\ldots,229)$ given by (3), $\hat{\underset{\sim}{\beta}}_j$ obtained from (2) and sorted; Fig. 3.k $(k = 1,\ldots,12)$ consists of plots for $j = k + 120,\ k + 132,\ldots,k + 228$ $(j \leq 229)$. (The darker curves denote these plots obtained in part from January 1971 or 1975.)



denoted by a thicker line. These are the ones whose coefficients are estimated from January 1958 or 1961. The plots in Fig. 3.k $(k = 1,...,12)$ are those obtained from (3) for $j = k + 120, k + 132,...,k + 228$ $(j \leq 229)$. The plots in Fig. 3.k are similar except for the ones denoted by a thicker line. These are the ones whose coefficients are estimated from January 1971 or 1975.

Thus Andrews' plots can be used as a graphical method not only to examine changes over time in the parameters but also to detect abrupt changes in the observations reflected by changes in the parameters of the model over time. As mentioned elsewhere, the Andrews' plots can also be used to determine the period length when this is unknown.

5. Acknowledgement

This work was done while the author was a visitor at the Mathematics Research Center at the University of Wisconsin-Madison. The author would like to thank the good offices of the Mathematics Research Center and in particular Mr. F. W. Sauer who did the computer graphics.

Bibliography

Andrews, D. F. (1972). Plots of high-dimensional data. Biometrics 28, 125-136.

Herzberg, A. M. (1981). Andrews' plots and their applications. University of Wisconsin-Madison, Mathematics Research Center, Technical Report 2203, and Proceedings of Third National Scientific Conference on Organization and Automation of Experimental Studies, Bulgaria.

Herzberg, A. M. and Hickie, J. S. (1981). An investigation of Andrews' plots to show time variation of model parameters. To appear.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 2239	2. GOVT ACCESSION NO. AD-A103	3. RECIPIENT'S CATALOG NUMBER 852
4. TITLE (and Subtitle) (6) AN EXAMPLE OF THE USE OF ANDREWS' PLOTS TO DETECT TIME VARIATIONS IN MODEL PARAMETERS AND OUTLYING OBSERVATIONS		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Agnes M. Herzberg (14) MRC-TSR-		8. CONTRACT OR GRANT NUMBER(s) (15) DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709	(11)	12. REPORT DATE July 1981
		13. NUMBER OF PAGES 8
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (9) T / 50		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited. (12)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Andrews' plots, time series, outliers, spurious observations, exploratory analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Andrews (1972) introduced a method of plotting high-dimensional data in two dimensions. This method is exploited as a graphical tool for the examination of changes over time in the parameters of a time series model. An example using a Fourier series model is given to illustrate the method. It is also shown how outlying observations in the data can be found.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 221200

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)